Dear Editor and reviewers:

We are submitting our revised manuscript, entitled "**Carbon emission reduction requires attention to the contribution of natural gas use: Combustion and leakage**" to *Atmospheric Chemistry and Physics*.

We thank the Associate Editor and reviewers for the detailed and helpful comments to improve the manuscript. Responses to the individual comments are provided below. Reviewer comments are in **bold**. Author responses are in blue plain text. Modifications to the manuscript (Tracked changes) are highlighted in red, similar issues are merged into one point, the numbering of the figure in this responses letter is the same as the manuscript or the supplementary.

The submitted manuscript has been revised based on reviewers' comments.

Sincerely,

Guiqian Tang,

Professor

Institute of Atmospheric Physics, Chinese Academy of Sciences

Beijing, China

Review of updated version of "Carbon emission reduction requires attention to the contribution of natural gas use: Combustion and leakage".

First and foremost, I want to congratulate the authors on working very hard and addressing each and every question I have raised. I believe that their effort has led to a more technically valid paper. I also believe that the paper now has stronger conclusions and deeper scientific merit. I have a few remaining questions, mostly about new points raised in their responses. I have structured some of my points to help the authors to write their "Caveats and Limitations" section of the conclusion (https://www.atmospheric-chemistry-and-physics.net/policies/guidelines_for_authors.html). I believe that with another round of much less intensive revisions, that the paper will be ready to contribute substantially to the ACP and wider academic community.

Response: We sincerely thank you for taking the time to review our revised manuscript and for your exceptionally positive and constructive feedback. We are greatly encouraged to learn that our efforts have strengthened the scientific rigor and conclusions of the paper. We fully agree with your assessment and are committed to addressing your final suggestions for enhancing the "Caveats and Limitations" section in the revisions. We have added the "Limitations" section as "4.4 section" and modified the corresponding part of the "Conclusions" in the revised manuscript.

In response to the specific points you raised, we will address each one and integrate them systematically. Once again, we deeply appreciate your expert guidance and sustained support throughout the review process.

(1) The author responds: "However, few studies have evaluated the impact of using different combination of time window and quantile on background value calculation. Yet, the choice of both the time window length and the quantile does indeed affect the final calculated background concentration. Here, using mobile measurement results near the gas storage tank in summer as an example, we evaluated the impact of different window-quantile combinations on background value calculation. This part has been added to supplementary."

The idea of using a background calculation is inherently complex because the

observations of both the background and the signal contain uncertainty. Even if you had access to the data of Shangdianzi, it would still contain uncertainty. For this reason, there are at least three recent papers which have worked to advance the idea of how to account for the join uncertainty's impact on emissions estimation, although none of them used observations from flux towers [Lu F. et al., 2025; Lu L. et al., 2025; Zheng et al., 2025]. I am not sure if you have the ability to repeat their approaches, or if you should think about how to write this new type of approach into the part of the conclusion that deals with caveats and limitations.

Response 1: Thank you very much for putting forward this crucial point of view, the sensitivity analysis we added (the selection of evaluation time window and quantile) only touched the surface of the uncertainty of background value estimation, but failed to solve the observation uncertainty of background value and enhanced concentration, and how they jointly spread and affect the uncertainty of final emission estimation.

Thank you again for pointing out the latest work of Lu et al. (2025, two papers) and Zheng et al. (2025). We have reviewed these literature, which jointly model the uncertainty of background concentration and observed concentration, and quantify their impact on the posterior distribution of emission flux through Bayesian or Monte Carlo methods. These methods indeed represent the forefront direction of methodology in this field. However, the main objective and data foundation of this study are more focused on quantifying the emission characteristics of potential natural gas leakage sources in Beijing using the observed methane concentration increments. Due to the limited data structure of the current observation scheme (such as the lack of multi-point synchronization and long-term stable background constraints), there are significant challenges in fully implementing the probability inversion framework described in Lu et al. (2025) or Zheng et al. (2025). Based on your suggestion, we will seriously and specifically elaborate on this limitation in the third paragraph of "4.4 Limitations section" as following:

When quantifying CH_4 leakage from different natural gas facilities, we adopted a quantile based deterministic method to separate background concentration from enhanced signals, and mainly explored the sensitivity brought by algorithm parameter selection. However, this framework has a fundamental limitation: it fails to incorporate the inherent observational uncertainty of background concentration and on-site observed concentration into a unified probabilistic analysis. The observation error of background value and enhanced signal are coupled (Lu et al., 2025; Zheng et al., 2025), they will propagate together, and significantly affect the final uncertainty interval of emission estimation. Our current sensitivity analysis can only be one step in such comprehensive uncertainty quantification work (i.e. identifying sensitivity to parameter selection), and future work should focus on: (1) systematically quantifying the instrument errors used in this study; (2) integrating these prior uncertainties into the inversion process of emissions using probabilistic frameworks such as Bayesian inference or error propagation models; (3) expanding the emission estimation from a single 'best estimate' to a probability distribution that includes confidence intervals.

(2) The author responds: "Replace identified outliers with linearly interpolated values from adjacent points. Consecutive outliers ≤ 3 are treated as a single outlier; consecutive outliers ≥ 4 are considered local trends and excluded from outlier classification."

I would think that you would need to carefully consider the footprint during these times before they are classified as either an outlier or a trend. The impact would be vastly different if the footprint during these times is similar to the normal conditions that you have demonstrated, or looked different from the normal conditions you have demonstrated. It also is possible that you are removing an actual emissions event, and replacing it with an interpolated value which is smaller, in which case you have just underestimated the actual emissions. Perhaps this is reasonable, but perhaps not, especially given the uncertainty in the observations themselves, as mentioned above in point 1.

Response 2: Thank you for your suggestion. We acknowledge that it is not reasonable to directly eliminate these outliers if they were attributed to specific emission incidents rather than measurement errors from the instrument. Therefore, following your suggestion, we first examined the number of outliers for CO_2 and CH_4. We found that compared to the total sample size, the number of outliers for CO_2 and CH_4 was very small, accounting for 0.7 % and 0.62 % of the total sample size, respectively. This indicates that the removal of outliers may not have a significant impact on the final flux results; Moreover, we also examined the flux footprints corresponding to these outliers and found that approximately 85 % of them fell within the 90 % contribution source area, indicating that the majority of outliers were caused by measurement errors rather than actual emission events. In summary, using interpolation to handle outlier data has little impact on the final flux.

(3) The author responds: “It can be seen that the source area covers the most urban area of Beijing. It basically covers the entire Fifth Ring Road area of Beijing but does not extend to other provinces, thus excluding long-range transport from other provinces.”

This map shows a footprint out to the 90th percentile. It would be interesting to see how far the remaining 10% of the footprint looks, especially if the fluxes computed during that remaining 10% fall outside of the range of the fluxes occurring within this 90%. Some simple graphs of the statistics of the computed fluxes (i.e., PDFs) occurring at within each ring, as well as outside of the outer ring should help us address this issue.

Response 3: Thank you for your comment. This is a very critical technical issue. In short, the "100% contribution source area" of a flux tower is infinite, which cannot be defined physically and has no practical significance. Due to the theory of atmospheric diffusion, the extremely small contribution far away from the tower always exists in theory, leading to the region boundary tending to be infinite.; The "cumulative contribution source area" is a practical concept that can be calculated and has clear physical meaning. It is the actual output of all models such as Kljun et al. It sets a contribution percentage and calculates the corresponding spatial range. The result is a limited, closed and quantifiable area, which has practical application value. The contribution of a point on the surface decays exponentially with the distance from the maximum contribution zone. Most of the signals come from the area near the tower, while the contribution from the remote area is very small and buried in the background noise of the instrument. Therefore, it may be better to define and calculate a region with dominant contribution (such as 90%) than to pursue the illusory "100%". In the study of flux source area, a unified cumulative contribution threshold is used to define the "effective source area", which provides a common benchmark for the comparison of results between different sites and objective criteria for evaluating data quality and discussing spatial representativeness.

(4) The author responds: "Second, the explanation lies in errors associated with the turbulent flux measurement system. This uncertainty is difficult to quantify because the sources of error are diverse, such as signal loss due to frequency attenuation in closed-path systems, the occurrence of negative values when real fluxes approach zero caused by the instrument's low signal-to-noise ratio, and the failure of the steady-state assumption underlying the eddy covariance method under conditions of weak turbulence. Unfortunately, no study can fully quantify the causes of negative values in flux observations currently, particularly over highly heterogeneous urban surfaces, where quantifying these uncertainties becomes especially challenging. Due to weaker turbulence development at night, flux measurement uncertainty increases, and the probability of observing negative fluxes is higher. Fluxes frequently fluctuate around zero during these periods.".

This is not the interpretation that I would make. I would follow your comment that the fluxes around zero during these periods are actually approximately zero, plus some amount of white noise from your instrument. Therefore, instead of considering both the positive and negative fluxes, instead you should assume all are a function of white noise, and ignore all of the both very small negative and very small positive fluxes, since white noise is just as equally positive as negative. This will then reduce your overall number of valid points, and likely lead to an increase in the overall emissions, although this will need to be carefully considered. This is a response similar to what the papers talked about in response to point (1). This may be too difficult to work out for this paper, but it should at least be pointed out as being important for future study and/or as a limitation of the current approach

Response 4: Thank you for pointing this out and for proposing this highly insightful and more rigorous methodological approach. We fully understand and agree with your interpretation of treating fluxes around zero as white noise, which is conceptually more precise than our initial explanation. The scheme you described—establishing a threshold based on the statistics of the instrument's white noise and discarding all fluxes (both slightly positive and slightly negative) within that threshold—is theoretically the ideal approach for handling low signal-to-noise conditions. This method effectively distinguishes genuine biophysical signals from instrument noise, resulting in a more robust dataset. However, as you rightly surmised, precisely defining this noise threshold for our specific instrument and site history requires additional, non-trivial analysis (e.g., testing under absolutely stable conditions or using high-frequency data statistics to quantify the instrument noise level). This falls outside the immediate scope of our current study. Nevertheless, we are convinced that the approach you propose is the correct way forward and that ignoring it would be a disservice to the manuscript's completeness. Therefore, we have fully adopted your suggestion and have incorporated this key methodological limitation and your proposed advanced solution as an important outlook for future work in the revised manuscript. We have added this in the first paragraph of "4.4 Limitations section" as following:

The flux discussed in this study is net flux, which means considering both positive flux and negative flux simultaneously. It should be noted that for flux values close to zero (particularly the negative values observed at night), we have retained all the data points without employing a filtering method based on the statistics of instrument white noise. Whereas a more rigorous approach would be to model these fluxes fluctuating around zero as white noise and establish a statistical significance threshold based on this. Discarding all values within this threshold (including slightly positive and slightly negative ones) could effectively reduce noise-induced bias, although at the cost of data coverage. The development and application of such objective, instrument-physics-based filtering criteria represent an important direction for future research to enhance the quality and reliability of flux data, particularly under low-turbulence conditions.

(5) The author responds: "Another a essential point is that CO has a long lifespan in the atmosphere, and it takes several tens of days to decay into CO_2 (Drummond

et al., 2009; Weinstock et al., 1969), so the impact of long-distance transmission of CO is relatively small."

I have checked these references and they both refer to modeling studies or global average values. Due to CO's very large variation in concentration, and that its lifetime is related to OH, which also varies by orders of magnitude, local lifetimes may also vary substantially. Recent observational papers using satellites and light-physical models in tandem, have demonstrated that in highly emitting regions that the lifetime of CO in the actual atmosphere is far shorter[Lin et al., 2020; Wang et al., 2021; Wang et al., 2025]. One such paper has specifically demonstrated that the production of CO_2 from CO is not insignificant in Shanxi, which is directly upwind of Beijing at least some fraction of the time[Li et al., 2025]. Again, this additional work may be too much of an extension, but it could be mentioned as a limitation.

Response 5: Thank you for this extremely valuable comment and for directing our attention to the studies by Lin et al. (2020), Wang et al. (2021), Wang et al. (2025) and Li et al. (2025). You are absolutely correct that our initial statement regarding CO's long lifetime of tens of days and the subsequent dismissal of its regional chemical impact was an oversimplification and did not reflect the current state of scientific understanding. As you aptly highlighted, CO removal is highly dependent on OH radicals, which exhibit tremendous spatiotemporal variability. The recent observational studies you cited compellingly demonstrate that in intense emission regions like the North China Plain, the local lifetime of CO can be substantially shorter due to active local photochemistry, and its rapid oxidation to CO_2 is non-negligible (Li et al., 2025). This is crucial for understanding regional air pollution and assessing the "chemical influence" on CO_2 sources of upwind source regions on downwind cities (e.g., Shanxi on Beijing). We have added a part explaining this in the second paragraph of "4.4 Limitations section" as following:

For the source analysis of CO_2 and CH_4, we did not consider the impact of long-distance transportation. However, this impact may not be completely ignored. For example, in the upwind area of Beijing, Shanxi Province is a high-intensity area of anthropogenic pollutant emissions, where the actual lifespan of local CO may be significantly shortened due to the influence of local OH concentration. The CO_2 produced by CO there is not insignificant(Li et al., 2025), which may also be one of the sources of local CO_2 in Beijing. Therefore, it is necessary to combine regional chemical transport models to more accurately quantify the impact of local chemical coupling in future flux research.

(6) The author responds: "Unfortunately, due to the lack of basic data from other cities or provinces, we are unable to measure its specific value accurately, a rough method was applied to estimated China's overall natural gas leakage rate based on existing reports and literature as follows, we have modified the Section 4.2 according to updated national leakage rate of natural gas.".

I would be very careful with this approach. For example, I have personally observed a very large amount of piped gas in use in Shanxi, much more so than you would realize based on the population, or other statistics used in your approach. Simple scaling approaches have been demonstrated to miss these substantial sources[Qin et al. 2023; Hu et al., 2024]. Again, such work would not likely be scalable based on your techniques used, but at least such a mention should be put into the "Caveats and Limitations" section of the conclusion.

Response 6: Thank you for this crucial comment and for directing our attention to the relevant studies by Qin et al. (2023) and Hu et al. (2024). We fully agree with your assessment. Your specific example regarding piped gas use in Shanxi Province and the risk of simple scaling approaches missing such significant sources is absolutely valid. Upon careful review of the suggested literature, we fully appreciate that national-scale extrapolation based on population or macro-statistics is indeed unable to capture the substantial regional disparities in energy mix, infrastructure level, and industrial activity intensity, leading to significant uncertainties in the estimates. We acknowledge that, within the current framework of our study, we lack the comprehensive baseline data required to perform a more refined (e.g., facility-based or province-level) estimation. Therefore, the rough scaling method we employed serves as a expedient solution under data constraints. Following your suggestion, we have made a significant revision to the manuscript. We have added a new part in the last paragraph of “4.4 Limitations section” and modified the corresponding part in the Conclusions to explicitly highlight this key methodological limitation. The added text reads:

Although the national-scale extrapolation of the natural gas leakage rate conducted in this study carries substantial uncertainty. Our approach, which relied on a simplified scaling method due to data availability constraints, may fail to account for strong regional heterogeneities in natural gas consumption patterns and infrastructure conditions. As highlighted by recent literatures (Qin et al., 2023; Hu et al., 2024), such scaling methods can systematically miss substantial emission sources in specific regions (e.g., industrial hubs like Shanxi Province). Therefore, our national estimate should be interpreted as a rough attempt.

Future work should prioritize the development of more granular, bottom-up inventories based on province-level activity data and infrastructure surveys to achieve a more accurate and robust assessment of China's overall natural gas leakage.